CAI CENTER

TECH MEMO

THE DEVELOPMENT OF AN ON-LINE SEARCHED COORDINATE INDEX FOR USE IN TEACHING AND RESEARCH

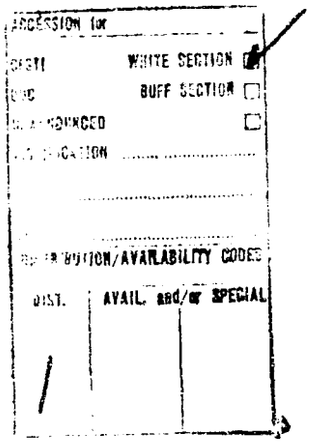
G. Jahoda and Ferol A. Foos
School of Library Science

Tech Memo No. 22
September 30, 1970

Project NR 154-280
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Washington, D. C.
Contract No. N00014-68-A-0494

FLORIDA STATE UNIVERSITY

40



TECH MEMO SERIES

The FSU-CAI Center Tech Memo Series is intended to provide communication to other colleagues and interested professionals who are actively utilizing computers in their research. The rationale for the Tech Memo Series is threefold. First, pilot studies that show great promise and will eventuate in research reports can be given a quick distribution. Secondly, speeches given at professional meetings can be distributed for broad review and reaction. Third, the Tech Memo Series provides for distribution of pre-publication copies of research and implementation studies that after proper technical review will ultimately be found in professional journals.

In terms of substance, these reports will be concise, descriptive, and exploratory in nature. While cast within a CAI research model, a number of the reports will deal with technical implementation topics related to computers and their language or operating systems. Thus, we here at FSU trust this Tech Memo Series will serve a useful service and communication for other workers in the area of computers and education. Any comments to the authors can be forwarded via the Florida State University CAI Center.

Duncan N. Hansen
Director

Computer Assisted Instruction Center

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Florida State University
Computer-Assisted Instruction Center
Tallahassee, Florida 32306

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

The Development of an On-line Searched Coordinate Index For Use in Teaching and Research

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Memo No. 22, September 30, 1970

5. AUTHOR(S) (First name, middle initial, last name)

G. Jahoda and Ferol A. Foos

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 30, 1970	32	4

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-68-A-0494	
b. PROJECT NO. NR 154-280	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
c.	
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Personnel & Training Research & Program Office of Naval Research Washington, D.C.

13. ABSTRACT

A model was developed for teaching coordinate index searching and preparation as well as for determining the effect of index and question variables on index performance. In this model coordinate index searching and preparation are considered as a series of decision-making steps. A coordinate index was prepared to 710 documents on library automation, systems studies in libraries, and indexing. The coordinate index has elements of vocabulary control but does not use roles, links, or weighting of index terms. Coordinate index searching and preparation were taught to library school students using classroom instruction, computer-aided instruction and on-line searching of test questions.

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

FORM 1473 (BACK)
DD 1 NOV 65
S/N 0101-807-6821

Security Classification
A-31409

THE DEVELOPMENT OF AN ON-LINE SEARCHED COORDINATE INDEX FOR USE IN TEACHING AND RESEARCH

G. Jahoda and Ferol A. Foos
School of Library Science

Tech Memo No. 22
September 30, 1970

Project NR 154-280
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Washington, D. C.
Contract No. N00014-68-A-0494

THE DEVELOPMENT OF AN ON-LINE SEARCHED COORDINATE INDEX FOR USE IN TEACHING AND RESEARCH

ABSTRACT

A model was developed for teaching coordinate index searching and preparation as well as for determining the effect of index and question variables on index performance. In this model coordinate index searching and preparation are considered as a series of decision-making steps. A coordinate index was prepared to 710 documents on library automatio systems studies in libraries, and indexing. The coordinate index has elements of vocabulary control but does not use roles, links, or weighting of index terms. Coordinate index searching and preparation were taught to library school students using classroom instruction, computer-aided instruction and on-line searching of test questions.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

This is a progress report on the development, preparation and initial use of an on-line searched coordinate index.[1] The report covers work done from October 1969 (the project's starting date) to June, 1970. The principal objectives of the project are the development of a model of an on-line searched index and its use for teaching coordinate index searching and preparation as well as for research. The project is planned in several stages:

1. Planning of index
2. Preparation of the first version of index and its use in teaching
3. Enlargement and refinement of index
4. Use of the enlarged and revised index in teaching and research.

The report includes work completed on stages 1 and 2 and plans for stages 3 to 4. Coordinate index searching and preparation was taught in several phases: classroom instruction, computer-aided instruction, on-line searching of instructor-generated and student-generated questions. The index now consists of 710 documents on library automation, systems studies in libraries and indexing and is being refined and enlarged this summer.

---

[1]This work was made possible through funds from the Office of Naval Research. We are grateful for this support, and for the programming work done by Tom McMurchie.

The equipment utilized for the project, available at the Computer-Assisted Instruction Center, includes an IBM 1500 Instructional System consisting of an 1800 central processor with 32,000 words (16 bits) of core, a 1502 station controller, sixteen 1510 CRT displays each with a keyboard and a light pen, one 1518 typewriter, and 5 2310 disk drives with removable disk packs of 512,000 16-bit words (1.024 million bytes). Additional perpherials include two 2401 tape units, one 1442 card reader/punch, and one 1443 line printer. In addition, the Center has interfaced a Digital Equipment Corporation PDP-8 to the 1500 system. This provides the ability to drive sixteen additional remote or local terminals.

II COORDINATE INDEX SEARCHING AND PREPARATION VIEWED AS DECISION MAKING STEPS

A model for teaching coordinate index searching and preparation was developed. In this model, an indexing system is viewed as a series of interacting processes, a suggestion made by Saracevic and Rees (1). Korotkin, et al. describe indexing as a series of decision-making processes (2). Snyder, et al. suggest in their report on methodologies of index system evaluation that the performance of each step in index preparation and searching needs to be examined since each step may have a contaminating effect on the overall performance of an index (3). The model of an indexing system given as Figure 1 was shaped by the views of the above cited researchers.

Each of the steps enclosed in solid lines in Figure 1 is considered decision-making step. The other steps, enclosed in dotted lines, are non-decision-making steps of a mechanical or clerical nature. Steps in coordinate index searching will be discussed first. The searcher begins by familiarizing himself with the question. At some point he has to decide that he either understands the question and can continue the process or does not understand the question and needs further help. If he understands the question (or at least thinks he does) his next decision is either to use the index or not to use the index for searching the question. The decision is based on the searcher's understanding of the question and the index (both its content and organization). If he decided to use the index for the question, he must also decide on what

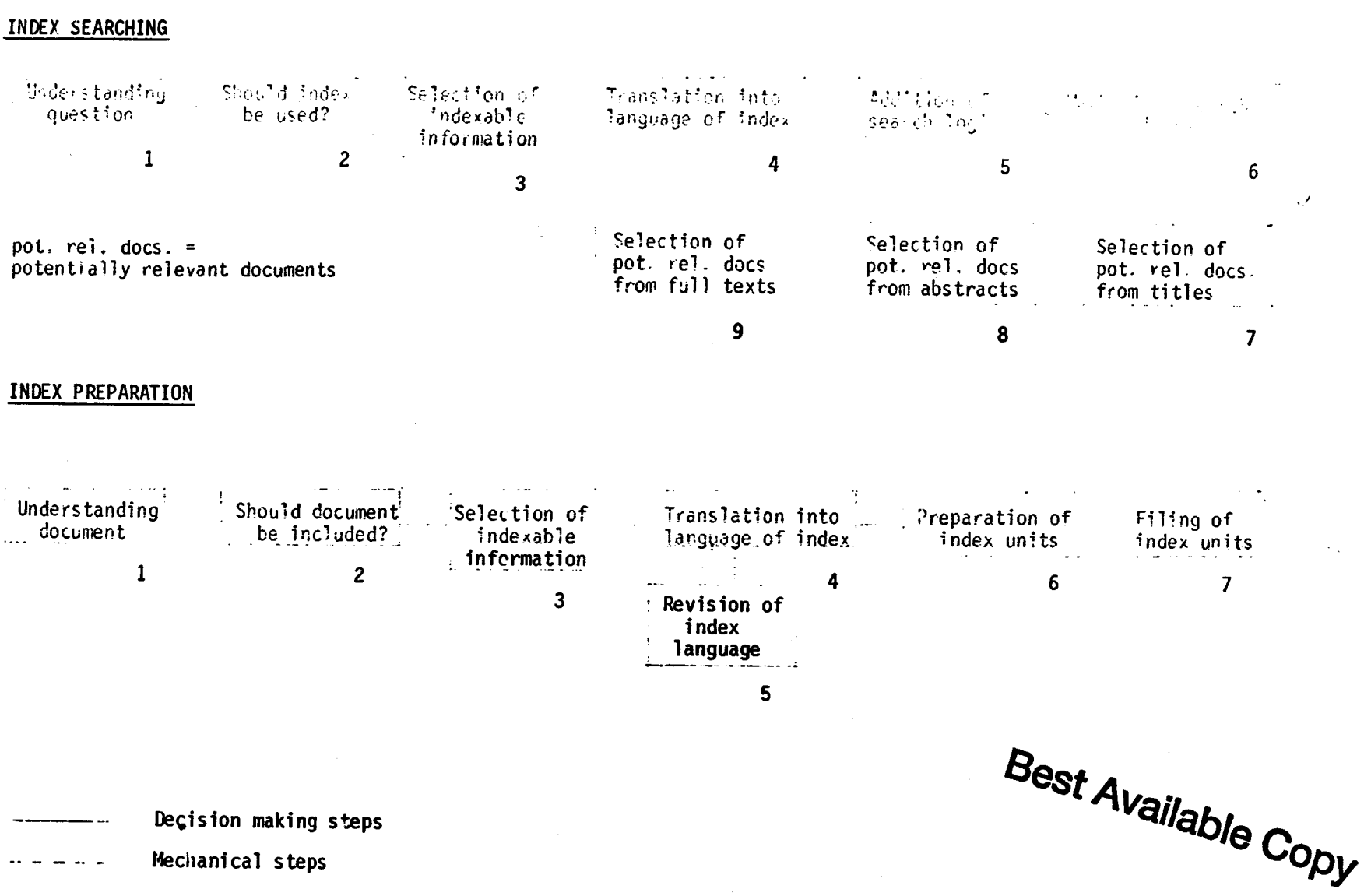


Figure 1.--Coordinate Index Searching and Preparation as Decision Making Steps

indexable information to select from the question--the words and phrases that carry the message of the question. The indexable information is next translated into the language of the index, a step considered a decision-making step though it could be a mechanical translation step if index language equivalents were available for all units of indexable information, something that is usually not so. The searcher then decides on logical connectors between indexing terms (descriptors). The descriptors and logical connectors are fed into the search mechanism for the next mechanical and non-decision-making step. The mechanical search step produces a list of documents that meet the search specifications. The searcher makes one or more decisions for each of the documents yielded by the index. He decides on the document's relevance or possible relevance based on document surrogates (title or abstract) and/or partial or complete document text.

There are similar decision-making and mechanical steps in coordinate index preparation. The indexer, like the searcher, begins by familiarizing himself with the document that is a candidate for inclusion in the index. After he decides that he understands the document (or does not understand it and makes a decision not depicted in Figure 1), the indexer either selects or rejects the document for the index. This decision is based on his understanding of the content of the document and rules for including as well as excluding documents from the index. Next, he selects indexable information by deciding on words and phrases which characterize the message of the document. The words and phrases are translated into descriptors as in the corresponding steps of index

searching. Indexable information not previously encountered in th index requires a decision to add either new descriptors, or to expand the meaning of existing descriptors, or perhaps to change one's mind about the necessity of a given unit of indexable information. The concluding steps in index preparation are the mechanical steps of preparing the index units (in one of several possible forms) and the filing of the indexed units, again in one of several possible ways.

I. PREPARATION OF INDEX

Steps in planning and preparing the index will now be summarized. Library automation, systems studies in libraries and (subsequently) indexing systems were selected as subjects to be included in the index. The choice was based on the background of the anticipated users of the index, in this case, library school students. When work was started on this project in October 1969, unsuccessful attempts were made to secure an index in machinable form to a collection of documents on the selected subjects. Existing subject authority lists did not appear to meet our needs. Several decisions had to be made before indexing could begin. It was decided to index documents on the basis of abstracts rather than full text, partly to expedite the preparation of the index, and partly because the literature on this subject provides no clear evidence that indexing from full text results in a better product than one obtained by indexing from abstracts. (This is perhaps because we lack measuring sticks for determining the adequacy of indexes, a topic that we propose to study during later stages of the project.) An examination of abstract bulletins covering the selected subjects led to the conclusion that Library and Information Science Abstracts would yield the largest number of documents; therefore, it was chosen as the first source of documents for the index. A coordinate index was selected as the type of index to use. Our choice was based on several factors, the

flexibility of this type of index in terms of possibility to combine index units at the time of the search of the index, the search capability of the available computer to take advantage of this search flexibility, and our desire to teach students coordinate index searching and preparation. Elements of vocabulary control were included both as a teaching device and as a variable to be tested in the research phase of the project. Decisions on coordinate index variables are discussed next.

A. Coordinate Index Variables. A coordinate index without roles or links was selected, again for more than one reason. The anticipated maximum size of the index (about 2,000 documents) is sufficiently small not to require precision devices. Also, there is no convincing evidence in the literature that roles and links are worth the additional indexing effort for collections of any size. Descriptors are to be either single words or phrases Indexable information selected for each document is to be as specific as possible. The selection of indexable information is to be done with the aid of a set of guidelines given in Figure 2.

B. Preparation of Subject Authority List. The indexing vocabulary is to be controlled with the aid of a subject authority list. Preparation of this list was begun with indexable information from the first 300 documents. Each of the about 2,500 unique words or phrases selected from these 300 documents was typed on an index card along with a number identifying the document from which the term was selected.

I. General guidelines for selection of indexable information

a. Base selection of indexable information on work reported not work planned unless entire document deals with plans.
b. Do not select as indexable concepts information not related to library automation, systems studies in libraries, or indexing.
c. Select indexable information as specifically as discussion of topic permits.
d. Select indexable information on one level only.
e. Do not select incidental information (minor matters mentioned in passing only).
f. Do not select name of meeting.

II. Questions intended to aid in the selection of indexable information

a. Are specific or types of libraries or information centers discussed?
b. Are library operations or services discussed?
c. Are products of library services or operations discussed?
d. Are these products evaluated?
e. What factors e g. speed, cost, are evaluated?
f. What techniques for evaluation are discussed?
g. Is equipment discussed?
h. Are equipment variables discussed?
i. Are other subjects dealing with library automation or systems studies or indexing discussed?
j. Is the type of article (review of the field, bibliography, survey, case history, philosophic discussion, research) useful in characterizing this document?
k. What is the date of the document?

Figure 2.--Guidelines for selection of indexable information.

The words and phrases were grouped into the following seven categories:

1. Abstract concepts
2. Equipment
3. Organizations
4. Products
5. Processes
6. Properties
7. Other

Terms in the larger categories were further subdivided until all groupings contained 200 or fewer terms. An example of subdivision of a category is the division of equipment terms into data processing equipment, data transmission equipment, photocopying equipment, and other equipment groupings. The groupings yielded a manageable number of terms for the next steps in the process, the selection of descriptors and "see" cross references, the establishment of specific generic relationships among descriptors in the same category, and the establishment of "see also" cross references among descriptors in different categories. Subject authority list decisions were recorded on the index cards. Two types of cards were prepared: "see" cross reference cards from synonyms to the chosen descriptor or combination of descriptors; descriptor cards consisting of the descriptor and its code (a one to three digit number), its more generic descriptor (one level higher only) and, for some but not all descriptors, scope notes, more specific descriptors (one lower level only), "see also" cross references, and "see from" cross references. A decision was made to list only one higher and one lower level of descriptors (out of a maximum of seven descriptor levels) to keep down the size of the subject authority list. Plans were to

provide a hierarchical arrangement of descriptors, but this was not ready for the initial use of the index. The first 300 documents yielded about [illegible] descriptors as well as several "anti-descriptors," concepts too broad to be useful, e.g. "Effect." Rules for using the subject authority list are given in its introduction, reproduced in the Appendix. The subject authority list was used to translate indexable information into descriptors for the first 300 documents and for 250 additional documents. Since multiple copies of the subject authority list were required and revision of the list was anticipated when additional documents are entered in the system, a master of the subject authority list was prepared on a magnetic tape typewriter. Multiple copies of the list were prepared from ditto masters.

C Encoding of Indexed Documents. A record of each indexed document is kept on a worksheet, illustrated in Figure 3. Worksheets for the 550 documents were taken to the computer terminal for encoding. For each document, the following information was put into machineable form at the computer console: document number, bibliographic citation of document in natural language, and codes for descriptors. The information keyed in at the terminal appeared simultaneously on the cathode ray tube. Proofing consisted of checking the copy on the worksheet against the cathode ray tube copy. Encoding was done at the rate of 12-20 documents per hours, depending on the speed and experience of the operator.

D Abstract Bulletin. Multiple copies of an abstract bulletin were prepared. The bulletin contains abstracts of the 550 documents included in the index and 160 documents to be indexed by students.

Document Number 491

491 The automation of information storage and retrieval systems, ... in Tague APLA Bull., 31(1) February 1967, 19-27. Bibliog.

Developments in computer technology mean that a user will be able to sit at an individual input-output console and communicate directly with the machine. This capability, together with the development of random-access storage systems, has re-awakened interest in the automation of the core problem of library systems--the storage and subject searching of an index file. Four areas are reviewed: (i) representation and storage of file items in mechanized information storage and retrieval systems; (ii) procedures for the retrieval of relevant documents; (iii) automated content analysis; (iv) overall design of information retrieval and library systems.

F.McA.

Indexable Information	Descriptor	Code
console	consoles	139
random-access storage systems	random access memory	520
storage	machine searches	367
subject searching	files	723
index file	mechanization	386
information retrieval	library services and operations	349
content analysis	systems analysis	634
systems design	on-line	442
library systems	1967-68	716
on-line	storage of words in computer	617
1967	machine classification	365

Figure 3.--Index Worksheet

E. On-line Searching of Index Searching instructions are ... in Figure 4. The computer configuration is given in the Intro- ... After the searcher signs on with either a general number or unique ... (used when record by searcher is to be kept), he has the choice ... search modes: LOAD 3 to view the complete bibliographic ... on each document selected by the system or LOAD 4 to view ... document number of each document selected by the system ... or console display of the two search modes are given in ... There is a third option, LOAD 5, differing from LOAD 4 only ... internal record keeping. Records of types of search logic are kept by the computer in the LOAD 5 search mode. Searches can be performed ... single descriptors or any logical AND, OR, NOT combination of ... descriptors. There are no restrictions on the number of descriptors ... can be coordinated by the machine at one time. There is now a ... limitation of a maximum of 640 unique document numbers ... can be searched the computer at one time. Search requests are made by typing the descriptor numbers and logical connectors AND, OR, NOT statements with parentheses when necessary. The first computer response is a display on the cathode ray tube of the number of documents that meet the search specifications. The searcher then has the options of viewing the document citations one by one (LOAD 3) or document number (LOAD 4 or LOAD 5) or to reformulate the search if either too many or not enough documents meet the search specifications. Another display option, a printout from teletype, is also possible with the equipment available but this option has not as yet been tested.

1. Get attention of the computer: Depress ALT CODE key, and while holding it down, press the INDEX key. When you see the cursor that marks typing line --
2. Type your identification code:)_____
3. Enter it into the computer: Depress the RETURN key. When cursor reappears --
4. Type either one of the following:
   For a display of bibliographic citations, type:)LOAD 3
   For a display of document numbers, only, type:)LOAD 4
5. Begin search typing by pressing the RETURN key. When cursor reappears --
6. To let the computer know you want to enter a logical search statement, type the letter: I
7. Press RETURN key to enter it. When the message "ENTER LOGICAL SEARCH STATEMENT" appears on the screen --
8. Type your logical search statement which should consist of: descriptor numbers connected with "AND" "OR" "NOT" (parentheses, if needed)
9. After proofreading your typing, enter it by pressing RETURN key.
10. If in step #4 you chose)LOAD 3, the computer will indicate the number of documents that are potentially relevant to your search statement. You now must use the light pen to indicate your choice of:
    viewing the first document citation, or
    returning to step #8 in order to type another search statement.
    If you choose to view a document citation, you will use the light pen to make a choice at the end of each citation. Press the symbol next to the word you choose. The point of the pen contacts a coded area.
11. If in step #4 you chose)LOAD 4, the computer will indicate the number of documents that are potentially relevant to your search statement and allow you to indicate with the light pen your choice of viewing or printing the numbers of these documents.
12. When you do not wish to continue entering logical statements, to get off the computer type: QUIT
13. Enter it by pressing the RETURN key. When the cursor reappears --
14. Turn off
15. Press with RETURN key. The computer indicates time you have used.

****

TO ERASE ERRORS MADE IN TYPING: (This can be done only before pressing the RETURN key.)

A. To erase one character at a time: Press the ALT CODE key, and while holding it down, press the BACK SPACE key as many times as needed.
B. To erase an entire line: Press the ALT CODE key and the upshift

? ? ? OTHER PROBLEMS ? ? ?

Figure 4 --continued

1. Did you remember to use the special APL characters (rather than the Coursewriter ones) when you were typing?
2. Did you proofread your typing before pressing the RETURN key to enter it into the computer?
3. Perhaps you confused 1 and I, or 0 and O?
4. Did you remember to enter your typing by pressing the RETURN key?
5. Did you get an error message? Call for help, when needed.
6. Is nothing happening?
   a. Perhaps the computer needs time for a lengthy search--or to service orders to printers or terminals.
   b. Perhaps something is wrong with the system or your terminal. Are others able to use their terminals?

Sample Question: Find articles on catalogs in book or card form. The catalogs should be in public libraries and should not be prepared from MARC tapes

Step 1. Determine question understanding by characterizing document titles as relevant, possibly relevant or non-relevant

QUESTION ANALYSIS: Select Documents on:
Find articles on catalogs in book or card form. The catalogs should be in public libraries and should not be prepared from MARC tapes

Document Title: Costs of Original Library Catalog Cards Produced by Computer at the Detroit Public Library

Relevant? ☐ Yes ☐ No ☐ Can't Tell

CORRECT?

☐ CONTINUE

Step 3. Selection of indexable information.

INDEXABLE WORD OR PHRASE SELECTION:

Type the numbers for words and/or the numbers connected with hyphens for phrases that represent indexable information for questions

---

Figure 5 -Examples of Coursewriter and APL program frames

Figure 5.--continued

Question: Find articles (1) on catalogs (2) in book (3) or card (4) form (5). The catalogs should be in public (6) libraries (7) and should not be prepared (8) from MARC (9) tapes (10).

CORRECT ANSWER: 2 3-2 4-2 6-7 9-10

Continue

Step 4. Translation of indexable information into descriptors with the aid of the subject authority list

CHECK SUBJECT AUTHORITY LIST:

Type descriptor numbers in blank space

Indexable Information:	Descriptor No.
-book catalogs	
-card catalogs	
-catalogs	
-public libraries	
-MARC TAPES	

CORRECT ANSWER:

-book catalogs	68
-card catalogs	86
-catalogs	93
-public libraries	508
-MARC	376

Continue

Step 5. Addition of search logic.

SEARCH LOGIC:

Type the search request, placing "AND" "OR" "NOT" and parentheses where needed:

Question: Find articles on catalogs (93) in book (68) or card (86) form. The catalogs should be in public libraries (508) and should not be prepared from MARC (376) tapes

CORRECT ANSWER:
508 NOT 376 AND (93 OR 68 OR 86)

Continue

Figure 5.--continued

Step 6. Mechanical search of index

ENTER LOGICAL DESCRIPTION:

508 NOT 376 AND (93 OR 68 OR 86)

7 DOCUMENTS REMAIN TO BE DISPLAYED

[CONTINUE [STOP

SEARCH MODE - LOAD 3

ENTER LOGICAL DESCRIPTION:

508 NOT 376 AND (93 OR 68 OR 86)

DOCUMENT NO. 45

A COMPUTER-CONTROLLED CHARGING SYSTEM AT ESSENDON PUBLIC LIBRARY,
W L BROWN AUST LIBR J., 16(6) DECEMBER 1967, 231-239.

6 DOCUMENTS REMAIN TO BE DISPLAYED

[CONTINUE [STOP

SEARCH MODE - LOAD 4

ENTER LOGICAL DESCRIPTION:

508 NOT 376 AND (93 OR 68 OR 86)

45
139
425
305
420
426
522

0 DOCUMENTS REMAIN TO BE DISPLAYED

[CONTINUE [STOP

Figure 5.--Examples of Coursewriter and APL program frames

THE USE OF THE INDEX IN TEACHING

The model was used in two courses offered by the School of Library Science during the Spring Quarter of 1970. The abstracting and indexing class had an initial enrollment of 17 students and a final enrollment of 16 students. The introductory course in information science, a required course for all graduate students in the School of Library Science and a prerequisite for the abstracting and indexing course, had an initial enrollment of 41 students and a final enrollment of 40 students.

A. Index Searching and Preparation by Abstracting and Indexing Students. All of the students in this course had already been exposed to six classroom hours on indexes in the introductory information science course. During the first meeting on indexing (and the third week of a ten-week term) the material covered in the introductory course was reviewed. Special emphasis was placed on the imperfect nature of indexing, something that students seem to find hard to understand until they start indexing themselves. Also reviewed are the economics of indexing and the principles of coordinate indexing. The initial classroom work (about three hours of classtime) dealt with the decision-making steps in coordinate index searching and preparation. The topics were covered first as lecture-discussions and then as class exercises. The index searching exercise performed in class with pencil and paper was repeated for emphasis at the computer console. Material covered in class is summarized below:

1. Understanding of question. Students were given an exercise to test their understanding of sample questions. They were asked to characterize coined document titles as relevant, possibly relevant, or non-relevant to a given question.

2. Selection of indexable information from question. This step was introduced by a discussion of words and phrases that are always, sometimes, or never to be selected as indexable information when searching this index. The groups of terms listed in the rules for selecting indexable information were suggested as an aid. For terms never to be selected, an analogy was made with delete word lists in keyword from title indexes. Words and phrases that are sometimes, but not always, used as indexable information turned out to be most difficult to deal with. Guidelines for determining whether to use single words or phrases were also discussed. Phrases are to be used when the combination of words is frequently or always used in our area of interest, e.g., library orientation, or when the combination of words in the phrase has a unique meaning not obvious from the meaning of the individual words, e.g., current awareness.

3. Translation of indexable information into descriptors. Steps in the preparation of a subject authority list were reviewed and students were given a small list of terms to be converted into a mini-subject authority list. When this exercise was completed, students translated indexable information into descriptors for questions handled in the previous steps.

4. Search logic formulation. Practice was provided before students added logical connectors (and parentheses when necessary) for the questions handled in previous steps.

The material on coordinate index searching covered in the classroom corresponds to steps 1, 3, 4, and 5 of the coordinate index search steps in Figure 1. Step two of the search model was omitted since it was assumed that all questions discussed in class were to be searched in the index. At the completion of the three-hour classroom work, the students were taken to the Computer-Assisted Instruction Center (CAI) to conduct the entire search process at individual computer consoles. The students searched the following eight questions at the consoles in two sessions of about two hours each.

1. I am interested in journal articles on computerized circulation systems operating on-line in a library.
2. I need reports of systems studies in libraries of industrial organizations.
3. What is the cost of telefacsimile systems?
4. Any information on microform readers in libraries?
5. I am looking for articles on surveys of library networks in Canada.
6. I want information about catalogs in book or card form. The catalogs should be in public libraries and should not be prepared from MARC tapes.
7. Are there any articles on publications in the field of chemistry that can be searched by machine?
8. I need journal articles about IBM 360 computers used in libraries in the USA.

Search steps 1, 3, 4, and 5 are written in Coursewriter II language. Step 6 is written in APL language. Step 7, the selection of documents based on document titles, uses the three level relevance judgment--relevant, possibly relevant, and non-relevant--used in the test for question understanding. Step 8, relevance judgment based on abstracts for documents judged either relevant or possibly relevant in the previous step, also uses the same three level relevance judgment. It should be pointed out that the student obtains the correct answer for each previous step except for step 7, and thus begins each new step with the correct answer for the previous one. Examples of frames for the Coursewriter and APL programs for question 6 are given in Figure 5.

After the index searching exercise, index preparation was covered in the classroom. While index preparation was discussed in terms of the decision-making model no formal exercises for teaching indexing were used. There was no time for the preparation of such exercises. Instead, sample abstracts were indexed in class. Indexable information for six abstracts was selected by the students with the aid of the rules. Disagreements in the selection of indexable information were discussed. The indexable information for these six abstracts were translated into descriptors. Students were also introduced to the procedure of adding new descriptors into the system. Each student indexed ten documents that were included into the system. The 160 new documents brought the index to 710 documents. Student-suggested new descriptors were screened by the project staff but student indexing was only spotchecked and not edited systematically. The 160 new documents added to the index produced

a number of changes to the subject authority list These changes were typed and distributed as a supplement to the subject authority list since time did not permit the insertion of these changes in the master copy of the list.

The last phase of the class exercise consisted of test-searching of the index The exercise was intended to give students additional practice in index use and to illustrate reasons for non-retrieval of relevant documents Each student was asked to submit two or more test questions to be searched by other students Each question had to include at least one document indexed by the question formulator and had to require both "and" as well as "or" search logic Sixteen questions were selected for test searching with each student searching five questions other than his own The searchers submitted to the formulator a record of search logic used and a list of relevant as well as possibly relevant documents selected from the index The question originator first compared his search results against test searchers' results for the question, adjusting his search results if additional relevant or potentially relevant documents were selected by the test searchers Recall ratios for the questions were then determined The average recall ratio for the 16 questions (sum of average recall for individual questions divided by 16) was 75% Reasons for non-retrieval of relevant and possibly relevant documents included (in decreasing order of occurrence) difference in relevance judgment, incorrect search logic, clerical error in transcription or coding, too many descriptors used as logical products, use of generic instead of specific descriptor, failure to use generic and specific descriptors

B The Use of the Decision Making Model in the Introductory Information Science Course. The index was used in this course for demonstration purposes. Two topics discussed in the course, on-line searching of computers and coordinate indexes, were demonstrated with the aid of the index. Students spent about half an hour at the computer console learning how to sign on and how to perform simple coordinate index searches already formulated for them. The index was also used by a fourth of the class on a voluntary basis as an aid in selecting references for a take-home final examination. Students were asked to write about the implications of on-line searched indexes for libraries in general or for type of library. The ten students who made use of the on-line searched index were not given any incentive to do so. They selected the on-line searched index over more readily available printed indexes (the CAIS is about a ten minutes' walk from the School of Library Science, and the index could only be searched on Wednesday evenings or Saturday afternoons).

Preliminary results of the use of the decision-making model and the on-line searched computer index appear encouraging. Students in the abstracting and indexing class were asked to comment anonymously about the course and their response was quite favorable on the whole. There are a number of improvements to be made. More instruction time needs to be spent on selection of indexable information for both questions and documents, the preparation and use of the subject authority list, and the search logic formulation.

The Coursewriter program for steps 1, 3, 4, and 5 of index searching is now primarily a drilling device. The correct answer is given or reasons for considering answers incorrect are given. Expansion of the program with branching for incorrect answers and use of the program for teaching index preparation as well as index searching would be useful. The API program document title display option (LOAD 3) was too slow with six or more students working simultaneously. Load 4 was developed when this became known. Students returned to go from document number to document abstract for checking the title. The entire system had to be switched from the Coursewriter program to the API program during the index searching exercise. This required waiting until all of the students finished with the Coursewriter program before the system could be switched to API and about five minutes to switch from one system to another. This was considered only a minor inconvenience that did not appear to cause any problems.

The absence of a hierarchical list of descriptors caused some search problems since at least some of the students did not follow the specific-generic descriptor network in the alphabetically arranged subject authority list.

The supplement of the subject authority list was not consulted at all times and this also resulted in search failures. Other problems with the index and work now in process in solving these problems is discussed next.

WORK IN PROGRESS AND PLANNED

A Revision of Subject Authority List

The test searches have brought out a number of problems with the index vocabulary. These problems are now being examined

1. Specific-generic relationships among descriptors

The listing of only one higher and one lower level of descriptor per entry has resulted in a shorter subject authority list than would otherwise have been the case but has also caused some problems in searching. The planned hierarchical list of descriptors that will be a supplement to the alphabetic list of descriptors is intended to solve this problem

2. Inconsistent use of generic descriptors

The following four descriptors caused problems since they were used inconsistently both in indexing and in searching: libraries, computers, computerization, and U S. The rules for indexing have been changed to make these four descriptors implied if the concepts cannot be expressed more specifically. For example, an article dealing with university libraries will be indexed under university libraries but an article dealing with libraries in general will have the concept of libraries implied since most documents deal with libraries or specific types of libraries

3. Proper nouns

Rules for proper noun descriptors were reviewed. Proper noun descriptors are now used for computers, some unique pieces of equipment

e g , IBM 357 Data Collection System, library associations, research institutes, research projects, clearinghouses, information analysis centers, program languages, theories, selected geographic areas, and indexing systems.

4 Word variants

Inconsistencies in the use of word variants as descriptors is also being analyzed Word variant descriptors are now characterized as process, product, or property descriptors and guidelines for separating or consolidating word variants as descriptors are now being developed

5 Type of document descriptor

Definitions of "types of document" descriptors, e g. case histories, reviews, surveys, are being prepared or expanded to assist the indexer and searcher in the use of these terms.

6 Scope notes and cross-references

Students searching the index had problems in using some other descriptors The subject authority list is now being examined for needed scope notes and cross-references, again to assist both the indexer and searcher in making more consistent decisions

B Use of the Index in Instruction

The revised and enlarged index will be used in the two previously mentioned library school courses this fall A doctoral student in the School of Library Science has proposed the use of the index for her dissertation. She proposes a comparison of existing methods for teaching the six hour segment on indexing in the introductory information science

course with a computer-managed multi-instructional media technique. The experimental method is to use taped lectures, slides as well as computer-aided instruction. Plans are also to use the index as an aid in the preparation of term papers. In this application, students will use the index as an aid in the selection of termpaper topics and references.

The instructional material for the abstracting and indexing course will be revised in line with the findings from the preliminary test of the system. Computer-aided instructional material for coordinate index searching and preparation will be prepared and the section on test searching of the index will be expanded.

C. Use of the Index Model for Research

Plans are to determine the effect of several question and index variables on index performance. While overall and final index performance will be measured in terms of recall ratio, precision ratio, and search time, intermediate search results will also be observed. "Readings" will be taken during each of the decision-making steps in index searching to measure the effect of the manipulated variable at the first point where such a variable may have an influence on the output. This procedure was suggested by Snyder, et al. (4) as a means for determining which part of the system is behaving improperly.

REFERENCES

1. T. Saracevic, & A. M. Rees. Towards the Identification and Control of Variables in Information Retrieval Experimentation. J. Doc. 23:7-19 (1967).

2 Kortokin, A. L., et al. Indexing Aids, Procedures and Devices. Rome Air Development Center, RADC-TR-64-582. April, 1965 (AD 616 342), p. 57.

3. Snyder, M. B., et al. Methodology for Test and Evaluation of Document Retrieval Systems: A Critical Review and Recommendations. Human Sciences Research, Inc., January 1966.

4. Snyder, M. B., et al. op. cit p. 64

APPENDIX

SUBJECT AUTHORITY LIST

(Prepared September, 1970)

INTRODUCTION

The subject authority list (sal) is intended to assist the indexer and the index searcher in the selection of the correct indexing terms, called descriptors in this coordinate index. The sal is based on indexable information from 1866 documents on library automation, indexing and systems studies in libraries. It includes about 1000 descriptors and approximately 2000 cross-references. The structure of descriptors and cross-references is illustrated by the following entries:

EXAMPLES OF SAL ENTRIES

"See" reference - one to one relationship

Subroutines see Computer programs and systems 498

"See" reference - one to more than one relationship

Kilgour's truncation algorithm see Computer programs and systems 498 and Truncated 672

Descriptor

Computer programs and systems 498

Includes named indexes or data in machineable form with programs for searching	SCOPE NOTE
BT Information sources, machineable	BROADER (MORE GENERIC) DESCRIPTOR

NT	FRONCY, computer program	NARROWER (MORE SPECIFIC)
NT	TERPS	DESCRIPTOR
NT	ABACUS	NOTE: ONLY THREE LEVELS OF
NT	MEDLARS	DESCRIPTORS ARE GIVEN ON
NT	Citation Identifier	ANY DESCRIPTOR ENTRY. SINCE
NT	EDUCOM	THERE ARE UP TO SEVEN LEVELS
NT	MASTER CONTROL	OF SPECIFICITY IN THE INDEX,
NT	ALPHA	BT S AND NT'S HAVE TO BE
NT	Pattern Learning Parser	CHECKED FOR MORE SPECIFIC
NT	ATS	AND GENERIC DESCRIPTORS
X	Routines	
X	Subroutines	SEE FROM REFERENCES
X	Software	NOTE: FOR TERMS TRANSLATED
X	Package programs	INTO MORE THAN ONE DESCRIPTOR
X	Kilgour's truncation algorithm (Truncated)	THE SECOND AND SUBSEQUENT TERMS ARE PLACED IN PARENTHESES IN SEE FROM REFERENCES
X	Algorithmic programs	
SA	Program languages	SEE ALSO REFERENCES FOR RELATED DESCRIPTORS

Only one higher level (more generic) and one lower level (more specific) descriptor are given on any descriptor entry. If all specifically-generically related terms to a descriptor are sought, the entries of the more generic and more specific descriptors have to be checked.

There are several descriptors that have more than one number, e.g. Indexes = 276 or 911, and the date 1967-68 = 716 or 900 or 902 or 913. Since each descriptor number can only include 160 document numbers in the computer's memory, these descriptors have to be assigned more than one number. For indexing, use the latest (highest) number. For searching, formulate a logical sum statement containing each number used to characterize the descriptor.

ALPHABETIZATION

Alphabetization is word by word. Acronyms and initials are treated as words and interfiled in the alphabet. Punctuation marks are disregarded.

RULES FOR SPECIFIC GROUPS OF TERMS

Implied terms

These are generic descriptors that cover subjects so basic and universal in the index that they would not be useful in searching. They are assumed to be relevant to every document in the index. The terms are: Libraries, Computerization, Mechanization, Computers, Data processing equipment, Automation, Data processing and US The US descriptor has a selective use explained in the scope note For the terms Libraries, Computers, Data processing equipment and US, the narrower terms (NT's) should be used when applicable, even though the generic terms are only implied

Geographic terms

Geographic descriptors are used to characterize publications, academic institutions and libraries. States in the US, UK, Canada and International are used specifically. Non-US is used for all other geographic areas not within the US

Organizations

Library associations: Specific proper name is used.

Organizations other than libraries: Specific proper names are used for Research projects and Research institutes Generic descriptors, rather than proper names, are used for International organizations, Societies, Trade associations, Industries, Publications, Commercial services, Government agencies, and all other organizations.

Academic institutions

Library schools: Generic descriptor, Library schools, is used, coordinated with appropriate geographic descriptor.

All other academic institutions: Type of institution descriptor is coordinated with geographic descriptor.

Libraries

Types of library descriptors used are the narrower (NT) terms under Libraries. Academic libraries is further narrowed to three types of academic libraries. Special libraries is further narrowed to five types. Type of library descriptor is coordinated with geographic descriptor (for public, school, academic, government, and research institute libraries) and with subject descriptors when applicable.

Equipment

Specific descriptor is used for any type of equipment, components and parts, but a trade name is not used. Computers are an exception; the trade name is used for each computer.

Individuals

Individuals' names are not used. Instead, the issuing agency or employer is used.